1974 Metropolitan Appendix B: Source and Accuracy

Table of Contents

Albany-Schenectady-Troy 3

Anaheim-Santa Ana-Garden Grove 12

Boston 21

Dallas 30

Detroit 39

Ft. Worth 48

Los Angeles-Long Beach 57

Memphis 66

Minneapolis-St Paul 75

Newark 84

Orlando 93

Phoenix 102

Pittsburgh 111

Saginaw 120

Salt Lake City 129

Spokane 138

Tacoma 147

Washington 156

Wichita 165

Albany-Schenectady-Troy 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Albany-Schenectady-Troy, N.Y., SMSA, 4,940 sample housing units were eligible for interview. Of this number, 180 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 470 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample. The AHS sample for the Albany-Schenectady-Troy SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 180 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the over-

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

all estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with a more reliable estimate of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would in-

clude the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (−) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Albany-Schenectady-Troy SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	70	10,000	720
200	100	25,000	1,100
500	160	50,000	1,480
1,000	230	100,000	1,820
2,500	370	250,000	760
5,000	510		

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Albany-Schenectady-Troy SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	80	2,500	400
200	110	3,500	480
500	170	5,000	580
700	200	7,500	740
1,000	250	10,000	880
1,500	300	12,700	1,030

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 19,600 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 960. Consequently, the 68-percent confidence interval is from 18,640 to 20,560 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 18,060 to 21,140 housing units with 90 percent confidence; and that the average estimate lies within the interval from 17,680 to 21,520 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 19,600 specified owner-occupied housing units with two bedrooms, 3,500 or 17.9 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 3,500 is approximately 430. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 17.9 percent is approximately 2.0 percentage points:

$$2.0=(100)\left(\frac{3{,}500}{19{,}600}\right)\sqrt{\left(\frac{430}{3{,}500}\right)^2-\left(\frac{960}{19{,}600}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 15.9 to 19.9 percent; the 90-percent confidence interval is from 14.7 to 21.1 percent; and the 95-percent confidence interval is from 13.9 to 21.9 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 1,400 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999.

Thus the apparent difference between the number of 1974 owner-occupied units with 3 or more bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 2,100. The standard error of 3,500 is 430, as shown above. Table I also shows the standard error on an estimate of 1,400 to be approximately 270. Therefore, the standard error of the estimated difference of 2,100 is about

$$503=\sqrt{(430)^2 + (270)^2}$$

Consequently, the 68-percent confidence interval for the 2,100 difference is from 1,590 to 2,610 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 1,280 to 2,920 housing units, and the 95-percent confidence interval is from 1,080 to 3,120. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;
(2) add to and subtract from 50 percent, the standard error determined in step 1; and
(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $24,600 in 1974. The base of the distribution from which this median was determined is 19,600 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 19,600 is 2.7 percentage points:

$$2.7=(100)\left(\frac{9,800}{19,600}\right)\sqrt{\left(\frac{710}{9,800}\right)^2-\left(\frac{960}{19,600}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 44.6 to 55.4.
3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first four categories that 6,000 owner-occupied housing units with two bedrooms, or 30.6 percent, had a value less than $20,000; and an additional 4,000 owner-occupied units with two bedrooms, or 20.4 percent, had a value between $20,000 and $24,999; and an additional 7,000, or 35.7 percent, had a value between $25,000 and $34,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$20,000 + (\$25,000-\$20,000)\left(\frac{44.6-30.6}{20.4}\right) = \$23,400$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$35,000-\$25,000)\left(\frac{55.4-30.6-20.4}{35.7}\right) = \$26,200$$

Thus, the 95-percent confidence interval ranges from $23,400 to $26,200.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which

they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items.

The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 1,800 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 1,200 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Anaheim-Santa Ana-Garden Grove 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-28
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-30
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Anaheim-Santa Ana-Garden Grove, Calif., SMSA, 4,940 sample housing units were eligible for interview. Of this number, 180 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 150 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Anaheim-Santa Ana-Garden Grove, Calif., SMSA was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population (the 1970 census universe) and units constructed since the 1970 census (the new construction universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate used to select the AHS sample for the SMSA was chosen so that the desired designated sample size would result. Thus, the overall sampling rate was about the same for the sample selected from both the central city and balance of this SMSA, since the sample was distributed proportionately between the central city and balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000–$5,999 . . .		
$6,000–$9,999 . . .		
$10,000–$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the 1970 census universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties, Part 1.**

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 180 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the 1970 census universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), and one noninterview cell for new construction sample housing units.

The first-stage ratio estimation procedure was employed for all sample housing units from the 1970 census universe. This factor was computed separately for all sample housing units within each 1970 census universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from 1970 census universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the 1970 census universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection

within each SMSA, units already selected for other Census Bureau surveys were deleted from the 1970 census universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with a more reliable estimate of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula under-

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Anaheim-Santa Ana-Garden Grove SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
200	160	25,000	1,730
500	250	50,000	2,400
1,000	350	100,000	3,230
2,500	560	250,000	4,260
5,000	790	500,000	3,120
10,000	1,110		

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Anaheim-Santa Ana-Garden Grove SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	100	1,500	410
200	150	2,500	540
500	240	3,500	650
700	280	5,000	790
1,000	340	6,000	870

estimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 68,300 specified owner-occupied housing units with five rooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,700. Consequently, the 68-percent confidence interval is from 65,600 to 71,000 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with five rooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 63,980 to 72,620 housing units with 90 percent confidence; and that the average estimate lies within the interval from 62,900 to 73,700 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 68,300 specified owner-occupied housing units with five rooms, 1,300, or 1.9 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 1,300 is approximately 390. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 1.9 percent is approximately 0.6 percentage points:

$$0.6=(100)\left(\frac{1,300}{68,300}\right)\sqrt{\left(\frac{390}{1,300}\right)^2-\left(\frac{2,700}{68,300}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 1.3 to 2.5 percent; the 90-percent confidence interval is from 0.9 to 2.9 percent; and the 95-percent confidence interval is from 0.7 to 3.1 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 200 specified owner-occupied units with five rooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with five rooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 1,100. The standard error of 1,300 is 390 as shown above. Table I also shows the standard error on an estimate of 200 to be approximately 160. Therefore, the standard error of the estimated difference of 1,100 is about

$$420=\sqrt{(390)^2+(160)^2}$$

Consequently, the 68-percent confidence interval for the 1,100 difference is from 680 to 1,520 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 430 to 1,770 housing units, and the 95-percent confidence interval is from 260 to 1,940. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with 5 rooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with five rooms was $33,200 in 1974. The base of the distribution from which this median was determined is 68,300 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 68,300 is 2.1 percentage points:

$$2.1=(100)\left(\frac{34,150}{68,300}\right)\sqrt{\left(\frac{1,980}{34,150}\right)^2-\left(\frac{2,700}{68,300}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 45.8 to 54.2

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first five categories that 6,700 owner-occupied housing units with five rooms, or 9.8 percent, had a value less than $25,000; and an additional 33,400, or 48.9 percent, had a value between $25,000 and $34,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$10,000)\left(\frac{45.8-9.8}{48.9}\right) = \$32,400$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$10,000)\left(\frac{54.2-9.8}{48.9}\right) = \$34,100$$

Thus, the 95-percent confidence interval ranges from $32,400 to $34,100.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.

2. Record check.—The comparison of census data with data obtained from an independent record source.

3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinter-**

views. Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional under-enumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.
3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970

census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 12,400 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 7,200 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Boston 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-28
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-30
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Boston, Mass., SMSA, 14,400 sample housing units were eligible for interview. Of this number, 740 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 1,450 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Boston, Mass., SMSA was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population (the 1970 census universe) and units constructed since the 1970 census (the new construction universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate used to select the AHS sample for the SMSA was chosen so that the desired designated sample size would result. The overall sampling rate for the SMSA did differ by central city and balance, since the sample for this SMSA was split equally between the central city and the balance of the SMSA.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the 1970 census universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 740 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within this SMSA, the factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the 1970 census universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), and one noninterview cell for new construction sample housing units.

The first-stage ratio estimation procedure was employed for all sample housing units from the 1970 census universe. This factor was computed separately for all sample housing units with each 1970 census universe noninterview cell mentioned above. The ratio estimate factor for each cell was equal to:

$$\frac{\text{1970 census count of housing units from 1970 census universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in

the sample selection for the 1970 census universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the 1970 census universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells	
Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census, based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census domolition data. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period, 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator

y = the denominator

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\,\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Boston, Mass., SMSA, for the Central Cities and for the Balance of the SMSA**

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central cities	Not in central cities
200	120	80	140
500	180	130	230
700	220	150	270
1,000 . .	260	180	320
2,500 . .	410	290	510
5,000 . .	580	410	710
7,500 . .	710	500	870
10,000 .	820	570	1,000
25,000 .	1,280	870	1,570
50,000 .	1,780	1,170	2,180
75,000 .	2,150	1,410	2,620
100,000	2,450	1,610	2,970
250,000	3,510	–	4,070
500,000	3,960	–	4,450
750,000	3,160		

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Boston, Mass., SMSA, for the Central Cities and for the Balance of the SMSA**

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central cities	Not in central cities
100	70	50	90
200	100	80	130
500	160	120	210
700	190	140	250
1,000 . .	230	170	300
2,500 . .	360	270	470
3,500 . .	430	330	560
5,000 . .	520	390	680
7,500 . .	640	490	850
10,000 .	740	570	1,000
15,000 .	920	720	–
25,000 .	1,220	–	–

Illustration of the use of the standard error tables.—Table A-2 in part C of this report shows that in this SMSA there were 67,000 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in column 1 of table I of this appendix shows that the standard error of an estimate of this size is approximately 2,030. Consequently, the 68-percent confidence interval is from 64,970 to 69,030 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 63,750 to 70,250 housing units with 90 percent confidence; and that the average estimate lies within the interval from 62,940 to 71,060 housing units with 95 percent confidence.

Table A-2 in part C also shows that of the 67,000 specified owner-occupied housing units with two bedrooms, 4,800, or 7.2 percent, were valued between $15,000 and $19,999. Column 1 of table I of this appendix shows that the standard error for 4,800 is approximately 570. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 7.2 percent is approximately 0.8 percentage points:

$$0.8=(100)\left(\frac{4,800}{67,000}\right)\sqrt{\left(\frac{570}{4,800}\right)^2-\left(\frac{2,030}{67,000}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 6.4 to 8.0 percent; the 90-percent confidence interval is from 5.9 to 8.5 percent; and the 95-percent confidence interval is from 5.6 to 8.8 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table A-2 in part C of this SMSA report shows that in 1974 there were 2,000 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 2,800. The standard error of 4,800 is 570 as shown above. Table I also shows the standard error on an estimate of 2,000 to be approximately 360. Therefore, the standard error of the estimated difference of 2,800 is about

$$670=\sqrt{(570)^2+(360)^2}$$

Consequently, the 68-percent confidence interval for the 2,800 difference is from 2,130 to 3,470 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 1,730 to 3,870 housing units, and the 95-percent confidence interval is from 1,460 to 4,140. Thus, we can conclude with 95-percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends

on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;
(2) add to and subtract from 50 percent, the standard error determined in step 1; and
(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table A-2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $31,700 in 1974. The base of the distribution from which this median was determined is 67,000 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 67,000 is 1.5 percentage points:

$$1.5=(100)\left(\frac{33,500}{67,000}\right)\sqrt{\left(\frac{1,450}{33,500}\right)^2-\left(\frac{2,030}{67,000}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.0 to 53.0.
3. From table A-2 in part C of this report it can be seen by cumulating the frequencies for the first five categories that 15,600 owner-occupied housing units with two bedrooms, or 23.3 percent, had a value less than $25,000; and an additional 26,700, owner-occupied housing units with two bedrooms, or 39.9 percent had a value between $25,000 and $34,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$10,000)\left(\frac{47.0-23.3}{39.9}\right) = \$30,900$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$10,000)\left(\frac{53.0-23.3}{39.9}\right) = \$32,400$$

Thus, the 95-percent confidence interval ranges from $30,900 to $32,400.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.
2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.
3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.

2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.
3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to complete reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed

that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 5,800 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 400 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Dallas 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Dallas, Texas, SMSA, 4,790 sample housing units were eligible for interview. Of this number, 120 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 320 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Dallas, Texas, SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 ...		
$6,000—$9,999 ...		
$10,000—$14,999 .		
$15,000 and over ..		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 120 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the over-

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

all estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would in-

clude the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Dallas SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
200	160	25,000	1,810
500	260	50,000	2,500
1,000	370	100,000	3,390
2,500	580	250,000	4,550
5,000	820	500,000	3,860
10,000	1,160		

TABLE II. **Standard Errors for Estimated Number of 1970-74 Lost Units for the Dallas SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	110	3,500	680
200	160	5,000	830
500	250	7,500	1,050
700	290	10,000	1,240
1,000	350	15,000	1,600
2,500	570	21,300	2,030

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 75,500 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,950. Consequently, the 68-percent confidence interval is from 72,550 to 78,450 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 70,780 to 80,220 housing units with 90 percent confidence; and that the average estimate lies within the interval from 69,600 to 81,400 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 75,500 specified owner-occupied housing units with two bedrooms, 13,700, or 18.1 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 13,700 is approximately 1,320. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 18.1 percent is approximately 1.6 percentage points:

$$1.6 = (100)\left(\frac{13{,}700}{75{,}500}\right)\sqrt{\left(\frac{1{,}320}{13{,}700}\right)^2 - \left(\frac{2{,}950}{75{,}500}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 16.5 to 17.9 percent; the 90-percent confidence interval is from 15.5 to 20.7 percent; and the 95-percent confidence interval is from 14.9 to 21.3 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 23,400 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999.

Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 9,700. The standard error of 13,700 is 1,320 as shown above. Table I also shows the standard error on an estimate of 23,400 to be approximately 1,740. Therefore, the standard error of the estimated difference of 9,700 is about

$$2{,}180 = \sqrt{(1{,}320)^2 + (1{,}740)^2}$$

Consequently, the 68-percent confidence interval for the 9,700 difference is from 7,520 to 11,880 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 6,210 to 13,190 housing units, and the 95 percent confidence interval is from 5,340 to 14,060. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $10,000 and $14,999 is greater than the number of units valued between $15,000 and $19,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $16,100 in 1974. The base of the distribution from which this median was determined is 75,500 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 75,500 is 2.1 percentage points:

$$2.1 = (100)\left(\frac{37{,}750}{75{,}500}\right)\sqrt{\left(\frac{2{,}160}{37{,}750}\right)^2 - \left(\frac{2{,}950}{75{,}500}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 45.8 to 54.2.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first two categories that 11,400 owner-occupied housing units with two bedrooms, or 15.1 percent, had a value less than $10,000; an additional 23,400 owner-occupied housing units with two bedrooms, or 31.0 percent had a value between $10,000 and $14,999; and an additional 13,700, or 18.1 percent, had a value between $15,000 and $19,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$(\$10{,}000) + (\$5{,}000)\left(\frac{45.8 - 15.1}{31.0}\right) = \$15{,}000$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$(\$15{,}000) + (\$5{,}000)\left(\frac{54.2 - 15.1 - 31.0}{18.1}\right) = \$17{,}200$$

Thus, the 95-percent confidence interval ranges from $15,000 to $17,200.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which

they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items.

The results of this study are presented in the census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.
3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 12,400 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 5,500 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Detroit 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-31
Illustration of the computation of the standard error of a difference App-31
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-32
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Detroit, Mich., SMSA, 14,310 sample housing units were eligible for interview. Of this number, 480 interviews were not obtained because, for occupied housing units, the occupants were, not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 900 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Detroit, Mich. SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate used to select the AHS sample for this SMSA was chosen so that the desired designated sample size would result. The overall sampling rate for this SMSA did differ by central city and balance, since the sample for this SMSA was split equally between the central city and the balance of the SMSA.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000–$5,999 . . .		
$6,000–$9,999 . . .		
$10,000–$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 480 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the over-

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

all estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would in-

clude the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\ (x/y)\ \sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Detroit, Mich., SMSA, for the Central Cities, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard Error		
	SMSA	In central cities	Not in central cities
200	150	120	160
500	230	190	250
700	280	230	300
1,000 ...	330	270	360
2,500 ...	520	430	560
5,000 ...	730	610	830
7,500 ...	900	750	1,060
10,000 ..	1,040	880	1,270
25,000 ..	1,630	1,460	1,760
50,000 ..	2,290	1,850	2,450
75,000 ..	2,770	2,250	2,960
100,000 .	3,170	2,570	3,370
250,000 .	4,720	3,830	4,830
500,000 .	5,930	-	6,050
750,000 .	6,200	-	6,350
1,000,000	5,680	-	-

TABLE II. Standard Errors for Estimated Number of 1970-1974 Lost Units for the Detroit, Mich. SMSA, for the Central Cities, and for the Balance of the SMSA

(68 chances out of 100)

Size of estimate	Standard Errors		
	SMSA	In central cities	Not in central cities
100	100	80	110
200	140	120	150
500	220	190	240
700	260	220	290
1,000 ...	300	270	340
2,500 ...	480	430	560
3,500 ...	570	510	670
5,000 ...	680	610	830
7,500 ...	840	750	1,060
10,000 ..	970	880	1,270
15,000 ..	1,200	1,090	1,670
25,000 ..	1,560	1,460	-
35,000 ..	1,870	1,790	-
57,500 ..	2,450	-	-

Illustration of the use of the standard error tables.—Table A-2 in part C of this report shows that in this SMSA there were 203,100 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in column 1 of table 1 of this appendix shows that the standard error of an estimate of this size is approximately 4,240. Consequently, the 68-percent confidence interval is from 198,860 to 207,340 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 196,320 to 209,880 housing units with 90 percent confidence; and that the average estimate lies within the the interval from 194,620 to 211,580 housing units with 95 percent confidence.

Table A-2 in part C also shows that of the 203,100 specified owner-occupied housing units with two bedrooms, 58,700, or 28.9 percent, were valued between $15,000 and $19,999. Column 1 of table I of this appendix shows that the standard error for 58,700 is approximately 2,460. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 28.9 percent is approximately 1.1 percentage points:

$$1.1 = (100)\left(\frac{58,700}{203,100}\right)\sqrt{\left(\frac{2,460}{58,700}\right)^2 - \left(\frac{4,240}{203,100}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 27.8 to 30.0 percent; the 90-percent confidence interval is from 27.1 to 30.7 percent; and the 95-percent confidence interval is from 26.7 to 31.1 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table A-2 in part C of this SMSA report shows that in 1974 there were 23,400 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 35,300. The standard error of 58,700 is 2,460 as shown above. Table I. also shows the standard error on an estimate of 23,400 to be approximately 1,570. Therefore, the standard error of the estimated difference of 35,300 is about

$$2{,}920 = \sqrt{(2{,}460)^2 + (1{,}570)^2}$$

Consequently, the 68-percent confidence interval for the 35,300 difference is from 32,380 to 38,220 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 30,630 to 39,970 housing units, and the 95 percent confidence interval is from 29,460 to 41,140. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table A-2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $21,000 in 1974. The base of the distribution from which this median was determined is 203,100 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 203,100 is 1.2 percentage points:

$$1.2 = (100)\left(\frac{101{,}550}{203{,}100}\right)\sqrt{\left(\frac{3{,}190}{101{,}550}\right)^2 - \left(\frac{4{,}240}{203{,}100}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.6 to 52.4.

3. From table A-2 in part C of this report it can be seen by cumulating the frequencies for the first four categories that 90,900 owner-occupied housing units with two bedrooms, or 44.8 percent, had a value less than $20,000; and an additional 52,200 owner-occupied housing units with two bedrooms, or 25.7 percent, had a value between $20,000 and $24,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$20{,}000 + (\$5{,}000)\left(\frac{47.6 - 44.8}{25.7}\right) = \$20{,}500$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$20{,}000 + (\$5{,}000)\left(\frac{52.4 - 44.8}{25.7}\right) = \$21{,}500$$

Thus, the 95-percent confidence interval ranges from $20,500 to $21,500.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources

of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.
2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.
3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing

units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the census memorandum, "Reinterview Results for Annual Housing Survey SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.
3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 12,000 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 4,200 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Ft. Worth 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Fort Worth, Tex., SMSA, 4,900 sample housing units were eligible for interview. Of this number, 140 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition, to units eligible for interview, 380 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Fort Worth, Tex., SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdication of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 140 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census, based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The quality or reliability of the independent estimate varied by SMSA, depending on

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

the completeness of the utility data or permit demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of total housing and an April 1970 independent estimate of total housing, generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period, 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a re-

sult, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Fort Worth, Tex., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	80	10,000	760
200	110	25,000	1,180
500	170	50,000	1,580
1,000	250	100,000	2,000
2,500	400	250,000	1,490
5,000	620		

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Fort Worth, Tex., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	70	1,500	290
200	100	2,500	400
500	160	3,500	490
700	190	5,000	620
1,000	230		

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 47,100 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 1,530. Consequently, the 68-percent confidence interval is from 45,570 to 48,630 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 44,650 to 49,550 housing units with 90 percent confidence; and that the average estimate lies within the interval from 44,040 to 50,160 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 47,100 specified owner-occupied housing units with two bedrooms, 8,600 or 18.3 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 8,600 is approximately 720. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 18.3 percent is approximately 1.4 percentage points:

$$1.4 = (100)\left(\frac{8,600}{47,100}\right)\sqrt{\left(\frac{720}{8,600}\right)^2 - \left(\frac{1,530}{47,100}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 16.9 to 19.7 percent; the 90-percent confidence interval is from 16.1 to 20.5 percent; and the 95-percent confidence interval is from 15.5 to 21.1 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 15,100 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 6,500. The standard error of 8,600 is 720 as shown above. Table I also shows the standard error on an estimate of 15,100 to be approximately 900. Therefore, the standard error of the estimated difference of 6,500 is about

$$1,150 = \sqrt{(720)^2 + (900)^2}$$

Consequently, the 68-percent confidence interval for the 6,500 difference is from 5,350 to 7,650 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 4,660 to 8,340 houisng units, and the 95-percent confidence interval is from 4,200 to 8,800. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $10,000 and $14,999 is greater than the number of units valued between $15,000 and $19,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $13,100 in 1974. The base of the distribution from which this median was determined is 47,100 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 47,100 is 1.8 percentage points:

$$1.8 = (100)\left(\frac{23{,}550}{47{,}100}\right)\sqrt{\left(\frac{1{,}140}{23{,}550}\right)^2 - \left(\frac{1{,}530}{47{,}100}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to an subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.4 to 53.6.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first two categories that 14,100 owner-occupied housing units with two bedrooms, or 29.9 percent, had a value less than $10,000; and an additional 15,100 owner-occupied housing units with two bedrooms, or 32.1 percent, had a value between $10,000 and $14,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$10{,}000 + (\$5{,}000)\left(\frac{46.4 - 29.9}{32.1}\right) = \$12{,}600$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$10{,}000 + (\$5{,}000)\left(\frac{53.6 - 29.9}{32.1}\right) = \$13{,}700$$

Thus, the 95-percent confidence interval ranges from $12,600 to $13,700.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency

in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 5,200 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 1,600 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Los Angeles-Long Beach 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-28
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Los Angeles-Long Beach, Calif., SMSA, 14,580 sample housing units were eligible for interview. Of this number, 740 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 700 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Los Angeles-Long Beach, Calif., SMSA was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population (the 1970 census universe) and units constructed since the 1970 census (the new construction universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate used to select the AHS sample for the SMSA was chosen so that the desired designated sample size would result. The overall sampling rate for the SMSA did differ by central cities and balance since the sample for this SMSA was split equally between the central cities and the balance of the SMSA.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000–$5,999 . . .		
$6,000–$9,999 . . .		
$10,000–$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe

were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the 1970 census universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 740 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the 1970 census universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated) and one noninterview cell for new construction sample housing units.

The first-stage ratio estimation procedure was employed for all sample housing units from the 1970 census universe. This factor was computed separately for all sample housing units within each 1970 census universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from 1970 census universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the 1970 census universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. How-

ever, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the 1970 census universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with a more reliable estimate of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

$$\frac{\text{Weighted count of matched lost units + Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Los Angeles-Long Beach, Calif., SMSA, for the Central Cities and for the Balance of the SMSA**

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central cities	Not in central cities
200.	200	190	200
500.	310	310	320
1,000 . . .	440	430	450
2,500 . . .	700	680	710
5,000 . . .	990	970	1,010
10,000 . .	1,400	1,360	1,430
25,000 . .	2,230	2,180	2,490
50,000 . .	3,310	3,230	3,380
100,000. .	4,350	4,240	4,440
250,000. .	6,670	6,510	6,810
500,000. .	8,940	8,720	9,120
1,000,000	11,110	10,840	11,340
1,500,000	11,440	—	—
2,000,000	10,080	—	—
2,500,000	6,020	—	—

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Los Angeles-Long Beach, Calif., SMSA, for the Central Cities, and the Balance of the SMSA**

(68 chances out of 100)

Size of estimate	Standard error		
	SMSA	In central cities	Not in central cities
200.	190	190	190
500.	300	300	290
700.	350	350	350
1,000 . . .	420	420	420
2,500 . . .	670	670	670
3,500 . . .	800	790	800
5,000 . . .	950	940	970
7,500 . . .	1,180	1,150	1,210
10,000 . .	1,360	1,330	1,420
15,000 . .	1,690	1,630	1,810
25,000 . .	2,230	2,110	2,490
35,000 . .	2,690	2,500	3,120
50,000 . .	3,310	—	—

when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

Illustration of the use of the standard error tables.—Table A-2 in part C of this report shows that in this SMSA there were 365,500 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in column 1 of table I of this appendix shows that the standard error of an estimate of this size is approximately 7,720. Consequently, the 68-percent confidence interval is from 357,780 to 373,220 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 353,150 to 377,850 housing units with 90 percent confidence; and that the average estimate lies within the interval from 350,060 to 380,940 housing units with 95 percent confidence.

Table A-2 in part C also shows that of the 365,500 specified owner-occupied housing units with two bedrooms, 48,700, or 13.3 percent, were valued between $15,000 and $19,999. Column 1 of table I of this appendix shows that the standard error for 48,700 is approximately 3,250. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 13.3 percent is approximately 0.8 percentage points:

$$0.8 = (100)\left(\frac{48,700}{365,500}\right)\sqrt{\left(\frac{3,250}{48,700}\right)^2 - \left(\frac{7,720}{365,500}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 12.5 to 14.1 percent; the 90-percent confidence interval is from 12.0 to 14.6 percent; and, the 95-percent confidence interval is from 11.7 to 14.9 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table A-2 in part C of this SMSA report shows that in 1974 there were 11,100 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 37,600. The standard error of 48,700 is 3,250 as shown above. Table I also shows the standard error on an estimate of 11,100 to be approximately 1,460. Therefore, the standard error of the estimated difference of 37,600 is about

$$3,560 = \sqrt{(3,250)^2 + (1,460)^2}$$

Consequently, the 68-percent confidence interval for the 37,600 difference is from 34,040 to 41,160 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 31,900 to 43,300 housing units, and the 95-percent confidence interval is from 30,480 to 44,720. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and

$14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table A-2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $28,100 in 1974. The base of the distribution from which this median was determined is 365,500 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 365,500 is 1.1 percentage points:

$$1.1 = (100)\left(\frac{182,750}{365,500}\right)\sqrt{\left(\frac{5,630}{182,750}\right)^2 - \left(\frac{7,720}{365,500}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.8 to 52.2.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first five categories that 140,500 owner-occupied housing units with two bedrooms, or 38.4 percent, had a value less than $25,000, and an additional 134,000 owner-occupied housing units with two bedrooms, or 36.7 percent, had a value between $25,000 and $34,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$10,000)\left(\frac{47.8 - 38.4}{36.7}\right) = \$27,600$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$10,000)\left(\frac{52.2 - 38.4}{36.7}\right) = \$28,800$$

Thus, the 95-percent confidence interval ranges from $27,600 to $28,800.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on

the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)

2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)

3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)

4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)

5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:

a. heating fuel
b. renters paying extra for utilities and/or fuel
c. bathtub or shower
d. flush toilet
e. telephone availability
f. year structure built
g. value of home
h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.
3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 16,600 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 4,300 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Memphis 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Memphis, Tenn.-Ark., SMSA, 4,660 sample housing units were eligible for interview. Of this number, 190 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 490 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Memphis SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 190 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the over-

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

all estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with a more reliable estimate of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would in-

clude the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2-\left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Memphis SMSA**

(68 changes out of 100)

Size of Estimate	Standard Error	Size of Estimate	Standard Error
200	110	10,000	780
500	180	25,000	1,200
1,000	250	50,000	1,610
2,500	390	100,000	2,020
5,000	560	250,000	1,400

TABLE II. **Standard Errors for Estimated Number of 1970 Units Lost Between 1970-1974 in the Memphis SMSA**

(68 changes out of 100)

Size of Estimate	Standard Error	Size of Estimate	Standard Error
100	80	2,500	410
200	120	5,000	580
500	180	10,000	820
1,000	260		

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 38,300 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 1,420. Consequently, the 68-percent confidence interval is from 36,880 to 39,720 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 36,030 to 40,570 housing units with 90 percent confidence; and that the average estimate lies within the interval from 35,460 to 41,140 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 38,300 specified owner-occupied housing units with two bedrooms, 10,600, or 27.7 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 10,600 is approximately 800. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 27.7 percent is approximately 1.8 percentage points:

$$1.8=(100)\left(\frac{10,600}{38,300}\right)\sqrt{\left(\frac{800}{10,600}\right)^2-\left(\frac{1,420}{38,300}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 25.9 to 29.5 percent; the 90-percent confidence interval is from 24.8 to 30.6 percent; and the 95-percent confidence interval is from 24.1 to 31.3 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 9,400 specified owner-occupied units with 3 or more bedrooms valued between $10,000 and $14,999.

Thus the apparent difference between the number of 1974 owner-occupied units with 3 or more bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 1,200. The standard error of 10,600 is 800, as shown above. Table I also shows the standard error on an estimate of 9,400 to be approximately 750. Therefore, the standard error of the estimated difference of 1,200 is about

$$1{,}100=\sqrt{(800)^2+(750)^2}$$

Consequently, the 68-percent confidence interval for the 1,200 difference is from 100 to 2,300 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from −560 to 2,960 housing units, and the 95-percent confidence interval is from −1,000 to 3,400. Thus, we cannot conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval includes zero or negative values. Nor can we conclude with 90 percent confidence that there is some evidence that the 10,600 is greater than the 9,400, since the 90-percent confidence interval also includes zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;
(2) add to and subtract from 50 percent, the standard error determined in step 1; and
(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $17,000 in 1974. The base of the distribution from which this median was determined is 38,300 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 38,300 is 2.0 percentage points:

$$2.0=(100)\left(\frac{19{,}150}{38{,}300}\right)\sqrt{\left(\frac{1{,}040}{19{,}150}\right)^2-\left(\frac{1{,}420}{38{,}300}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.0 to 54.0.
3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first three categories that 14,900 owner-occupied housing units with two bedrooms, or 38.9 percent, had a value less than $15,000; and an additional 10,600, or 27.7 percent, had a value between $15,000 and $19,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$15{,}000+(\$20{,}000-\$15{,}000)\left(\frac{46.0-38.9}{27.7}\right)=\$16{,}300$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$15{,}000+(\$20{,}000-\$15{,}000)\left(\frac{54.0-38.9}{27.7}\right)=\$17{,}700$$

Thus, the 95-percent confidence interval ranges from $16,300 to $17,700.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which

they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional under-enumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items.

The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "Overall, the reinterview results showed a fairly inconsistent response between the original and reinterview data."
2. "Our bias indicator, the net difference rate revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."
3. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checks, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 3,700 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 1,100 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Minneapolis-St Paul 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Minneapolis-St. Paul, Minn., SMSA, 4,720 sample housing units were eligible for interview. Of this number, 100 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 320 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Minneapolis-St. Paul, Minn., SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 100 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the over-

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

all estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with a more reliable estimate of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would in-

clude the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (−) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Minneapolis–St. Paul, Minn., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
200	170	25,000	1,820
500	260	50,000	2,520
1,000	370	100,000	3,410
2,500	590	250,000	4,590
5,000	830	500,000	3,900
10,000	1,170		

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Minneapolis-St. Paul, Minn., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	110	3,500	680
200	150	5,000	830
500	240	7,500	1,050
700	290	10,000	1,250
1,000	350	15,000	1,620
2,500	560	22,600	2,140

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 90,200 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 3,240. Consequently, the 68-percent confidence interval is from 86,960 to 93,440 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 85,020 to 95,380 housing units with 90 percent confidence; and that the average estimate lies within the interval from 35,460 to 41,140 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 90,200 specified owner-occupied housing units with two bedrooms, 13,700, or 15.2 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 13,700 is approximately 1,330. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 15.2 percent is approximately 1.4 percentage points:

$$1.4 = (100)\left(\frac{13{,}700}{90{,}200}\right) - \sqrt{\left(\frac{1{,}330}{13{,}700}\right)^2 - \left(\frac{3{,}240}{90{,}200}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 13.8 to 16.6 percent; the 90-percent confidence interval is from 13.0 to 17.4 percent; and the 95-percent confidence interval is from 12.4 to 18.0 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 6,000 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999.

Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 7,700. The standard error of 13,700 is 1,330, as shown above. Table I also shows the standard error on an estimate of 6,000 to be approximately 900. Therefore, the standard error of the estimated difference of 7,700 is about

$$1{,}610 = \sqrt{(1330)^2 + (900)^2}$$

Consequently, the 68-percent confidence interval for the 7,700 difference is from 6,090 to 9,310 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 5,120 to 10,280 housing units, and the 95-percent confidence interval is from 4,480 to 10,920. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $25,800 in 1974. The base of the distribution from which this median was determined is 90,200 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 90,200 is 1.9 percentage points:

$$1.9 = (100)\left(\frac{45{,}100}{90{,}200}\right)\sqrt{\left(\frac{2{,}380}{45{,}100}\right)^2 - \left(\frac{3{,}240}{90{,}200}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.2 to 53.8.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first four categories that 20,500 owner-occupied housing units with two bedrooms, or 22.7 percent, had a value less than $20,000; and an additional 22,100 owner-occupied housing units with two bedrooms, or 24.5 percent had a value between $20,000 and $24,999; and an additional 33,000, or 36.6 percent, had a value between $25,000 and $34,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$20{,}000 + (\$25{,}000 - \$20{,}000)\left(\frac{46.2 - 22.7}{24.5}\right) = \$24{,}800$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$25{,}000 + \left(\$35{,}000 - \$25{,}000\right)\left(\frac{53.8 - 22.7 - 24.5}{36.6}\right) = \$26{,}800$$

Thus, the 95-percent confidence interval ranges from $24,800 to $26,800.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which

they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items.

The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.
3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 10,100 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 2,900 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Newark 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-28
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-30
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Newark, N.J., SMSA, 4,820 sample housing units were eligible for interview. Of this number, 220 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 200 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Newark, N.J., SMSA was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population (the 1970 census universe) and units constructed since the 1970 census (the new construction universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the 1970 census. universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties, Part 1.**

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 220 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within this SMSA, the factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the 1970 census universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), and one noninterview cell for new construction sample housing units.

The first-stage ratio estimation procedure was employed for all sample housing units from the 1970 census universe. This factor was computed separately for all sample housing units with each 1970 census universe noninterview cell mentioned above. The ratio estimate factor for each cell was equal to:

$$\frac{\text{1970 census count of housing units from 1970 census universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the 1970 census universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the 1970 census universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells	
Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census, based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period, 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator

y = the denominator

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula under-

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Newark, N.J., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
200	160	25,000	1,770
500	260	50,000	2,450
1,000	360	100,000	3,310
2,500	570	250,000	4,390
5,000	810	500,000	3,410
10,000	1,140		

TABLE II. **Standard Errors for Estimated number of 1970-1974 Lost Units for the Newark, N.J.,SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	90	2,500	480
200	130	3,500	580
500	210	5,000	710
700	240	7,500	900
1,000	290	10,000	1,080
1,500	360	18,700	1,660

estimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 13,400 specified owner-occupied housing units with four rooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 1,280. Consequently, the 68-percent confidence interval is from 12,120 to 14,680 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with four rooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 11,350 to 15,450 housing units with 90 percent confidence; and that the average estimate lies within the interval from 10,840 to 15,960 housing units with 95 percent confidence.

Table 2 in part C also shows that of 13,400 specified owner-occupied housing units with four rooms, 800, or 6.0 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 800 is approximately 320. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 6.0 percent is approximately 2.3 percentage points:

$$2.3 = (100)\left(\frac{800}{13{,}400}\right)\sqrt{\left(\frac{320}{800}\right)^2 - \left(\frac{1{,}280}{13{,}400}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 3.7 to 8.3 percent; the 90-percent confidence interval is from 2.3 to 9.7 percent; and the 95-percent confidence interval is from 1.4 to 10.6 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 6,300 specified owner-occupied units with four rooms valued between $25,000 and $34,999. Thus the apparent difference between the number of 1974 owner-occupied units with four rooms valued between $25,000 and $34,999 and those valued between $15,000 and $19,999 is 5,500. The standard error of 800 is 320 as shown above. Table I also shows the standard error on an estimate of 6,300 to be approximately 900. Therefore, the standard error of the estimated difference of 5,500 is about

$$960 = \sqrt{(320)^2 + (900)^2}$$

Consequently, the 68-percent confidence interval for the 5,500 difference is from 4,540 to 6,460 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 3,960 to 7,040 housing units, and the 95-percent confidence interval is from 3,580 to 7,420. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with four rooms, valued between $25,000 and $34,999 is greater than the number of units valued between $15,000 and $19,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction

with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with four rooms was $32,800 in 1974. The base of the distribution from which this median was determined is 13,400 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 13,400 is 4.9 percentage points:

$$4.9 = (100)\left(\frac{6,700}{13,400}\right)\sqrt{\left(\frac{920}{6,700}\right)^2 - \left(\frac{1,280}{13,400}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 40.2 to 59.8.

3. From Table 2 in part C of this report it can be seen by cumulating the frequencies for the first five categories that 1,800 owner-occupied housing units with four rooms, or 13.4 percent, had a value less than $25,000; and an additional 6,300 owner-occupied housing units with four rooms, or 47.0 percent, had a value between $25,000 and $34,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$10,000)\left(\frac{40.2 - 13.4}{47.0}\right) = \$30,700$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$25,000 + (\$10,000)\left(\frac{59.8 - 13.4}{47.0}\right) = \$34,900$$

Thus, the 95-percent confidence interval ranges from $30,700 to $34,900.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.

2. Record check.—The comparison of census data with data obtained from an independent record source.

3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series

PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)

2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)

3. "In 1970, the definitional under-enumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)

4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)

5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:

a. heating fuel
b. renters paying extra for utilities and/or fuel
c. bathtub or shower
d. flush toilet
e. telephone availability
f. year structure built
g. value of home
h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.
3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that

units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 2,300 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 60 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Orlando 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Orlando, Fla., SMSA, 4,630 sample housing units were eligible for interview. Of this number, 70 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 470 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Orlando, Fla., SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 70 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the second-stage ratio estimation procedure, as well as the over-

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

all estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census, based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The quality or reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or permit demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of total housing and an April 1970 independent estimate of total housing, generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period, 1960-70).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error

above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the table below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory as well as estimates of characteristics of the 1970-74 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in table I.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory and for Estimated Number of 1970-1974 Lost Units for the Orlando, Fla., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	70	5,000	460
200	90	10,000	640
500	150	25,000	970
1,000	210	50,000	1,270
2,500	350	100,000	1,470

Illustration of the use of the standard error table.—Table 2 in part C of this report shows that in this SMSA there were 20,600 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 870. Consequently, the 68-percent confidence interval is from 19,730 to 21,470 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 19,210 to 21,990 housing units with 90 percent confidence; and that the average estimate lies within the interval from 18,860 to 22,340 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 20,600 specified owner-occupied housing units with two bedrooms, 5,100 or 24.8 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 5,100 is approximately 460. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 24.8 percent is approximately 2.0 percentage points:

$$2.0 = (100)\left(\frac{5,100}{20,600}\right)\sqrt{\left(\frac{460}{5,100}\right)^2 - \left(\frac{870}{20,600}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 22.8 to 26.8 percent; the 90-percent confidence interval is from 21.6 to 28.0 percent; and the 95-percent confidence interval is from 20.8 to 28.8 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 2,800 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999.

Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 2,300. The standard error of 5,100 is 460 as shown above. Table I also shows the standard error on an estimate of 2,800 to be approximately 360. Therefore, the standard error of the estimated difference of 2,300 is about

$$580 = \sqrt{(460)^2 + (360)^2}$$

Consequently, the 68-percent confidence interval for the 2,300 difference is from 1,720 to 2,880 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, to 90-percent confidence interval is from 1,370 to 3,230 housing units, and the 95-percent confidence interval is from 1,140 to 3,460. Thus, we can conclude with 95-percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $20,600 in 1974. The base of the distribution from which this median was determined is 20,600 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 20,600 is 2.3 percentage points:

$$2.3 = (100)\left(\frac{10,300}{20,600}\right)\sqrt{\left(\frac{650}{10,300}\right)^2 - \left(\frac{870}{20,600}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 45.4 to 54.6.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first three categories that 4,600 owner-occupied housing units with two bedrooms, or 24.8 percent, had a value less than $15,000; and an additional 5,100 owner-occupied housing units with two bedrooms, or 24.8 percent had a value between $15,000 and $19,999; and an additional 4,600, or 22.3 percent, had a value between $20,000 and $24,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{45.4 - 22.3}{24.8}\right) = \$19,700$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$20,000 + (\$5,000)\left(\frac{54.6 - 22.3 - 24.8}{22.3}\right) = \$21,700$$

Thus, the 95-percent confidence interval ranges from $19,700 to $21,700.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which

they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items.

The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.
3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 3,400 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 6,300 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Phoenix 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-28
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-30
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Phoenix, Ariz., SMSA, 4,870 sample housing units were eligible for interview. Of this number, 120 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 380 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Phoenix, Ariz. SMSA was selected from two sample frames— units enumerated in the 1970 Census of Housing and Population (the 1970 census universe) and units constructed since the 1970 census (the new construction universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the 1970 census universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 120 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the 1970 census universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated) and one noninterview cell for new construction sample housing units.

The first-stage ratio estimation procedure was employed for all sample housing units from the 1970 census universe. This factor was computed separately for all sample housing units within each 1970 census universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from 1970 census universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the 1970 census universe. Ordinarily, this would have been

controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the 1970 census universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula under-

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Phoenix, Ariz., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
200	140	10,000	970
500	220	25,000	1,510
1,000	310	50,000	2,070
2,500	490	100,000	2,750
5,000	690	250,000	3,330

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Phoenix, Ariz., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	80	5,000	610
300	150	7,500	750
500	190	10,000	880
700	220	15,000	1,100
1,000	270	20,000	1,300
2,500	420		

estimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (—) to plus (+).

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 52,600 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 2,110. Consequently, the 68-percent confidence interval is from 50,490 to 54,710 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 49,220 to 55,980 housing units with 90 percent confidence; and that the average estimate lies within the interval from 48,380 to 56,820 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 52,600 specified owner-occupied housing units with two bedrooms, 13,800, or 26.2 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 13,800 is approximately 1,110. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 26.2 percent is approximately 1.8 percentage points:

$$1.8 = (100)\left(\frac{13{,}800}{52{,}600}\right)\sqrt{\left(\frac{1{,}110}{13{,}800}\right)^2 - \left(\frac{2{,}110}{52{,}600}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 24.4 to 28.0 percent; the 90-percent confidence interval is from 23.3 to 29.1 percent; and the 95-percent confidence interval is from 22.6 to 29.8 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 4,500 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 9,300. The standard error of 13,800 is 1,110 as shown above. Table I also shows the standard error on an estimate of 4,500 to be approximately 650. Therefore, the standard error of the estimated difference of 9,300 is about

$$1{,}290 = \sqrt{(1{,}110)^2 + (650)^2}$$

Consequently, the 68-percent confidence interval for the 9,300 difference is from 8,010 to 10,590 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 7,240 to 11,360 housing units, and the 95-percent confidence interval is from 6,720 to 11,880. Thus, we can conclude with 95-percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard

error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $22,500 in 1974. The base of the distribution from which this median was determined is 52,600 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 52,600 is 2.1 percentage points:

$$2.1 = (100)\left(\frac{26,300}{52,600}\right)\sqrt{\left(\frac{1,540}{26,300}\right)^2 - \left(\frac{2,110}{52,600}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 45.8 to 54.2

3 From table 2 in part C of this report it can be seen by cumulating the frequencies for the first three categories that 20,800 owner-occupied housing units with two bedrooms, or 39.5 percent, had a value less than $20,000; and an additional 11,200 owner-occupied housing units with two bedrooms, or 21.3 percent, had a value between $20,000 and $24,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$20,000 + (\$5,000)\left(\frac{45.8 - 39.5}{21.3}\right) = \$21,500$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$20,000 + (\$5,000)\left(\frac{54.2 - 39.5}{21.3}\right) = \$23,500$$

Thus, the 95-percent confidence interval ranges from $21,500 to $23,500.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.

2. Record check.—The comparison of census data with data obtained from an independent record source.

3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series

PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)

2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)

3. "In 1970, the definitional under-enumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)

4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)

5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:

a. heating fuel
b. renters paying extra for utilities and/or fuel
c. bathtub or shower
d. flush toilet
e. telephone availability
f. year structure built
g. value of home
h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurment of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional

new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 6,700 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 19,500 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Pittsburgh 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Pittsburgh, Pa., SMSA, 4,920 sample housing units were eligible for interview. Of this number, 160 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 340 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Pittsburgh SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000–$5,999 . . .		
$6,000–$9,999 . . .		
$10,000–$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a three-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 160 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The third-stage ratio estimation procedure was employed for all sample units. This procedure was designed to adjust the AHS sample estimate of the October 1974 housing inventory for this SMSA to a final independent estimate of the October 1974 housing inventory. This ratio estimation factor equaled the following:

$$\frac{\text{Independent estimate of the October 1974 housing inventory for the SMSA}}{\text{AHS sample estimate of the housing inventory}}$$

The numerator of the ratio was derived by using 1970 census estimates of total housing units for the SMSA in conjunction with estimates of the change in the housing inventory since the census. These changes were based mostly on administrative records from utility companies, or, when utility data was not available, on estimates of new construction permits and post-census demolitions. The denominator of that ratio was obtained from the weighted estimate of AHS sample units, using the existing weight after the second-stage ratio estimation procedure. The computed third-stage ratio estimation factor was then applied to the existing weight for each sample unit and the resulting product was used as the final weight for tabulation purposes.

The effect of the third-stage ratio estimation procedure, as well as the over-

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

all estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with a more reliable estimate of the SMSA housing population.

In some of the Year I SMSA's, the third-stage ratio estimation procedure was not employed. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was employed because it was considered reliable (i.e., its relative bias was 3 percent or less over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would in-

clude the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2-\left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (—) to plus (+).

Table I. Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Pittsburgh SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
200	180	25,000	2,040
500	290	50,000	2,830
1,000	410	100,000	3,880
2,500	650	250,000	5,460
5,000	920	500,000	5,790
10,000	1,300	750,000	3,360

Table II. Standard Errors for Estimated Number of 1970-1974 Lost Units for the Pittsburgh SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	120	5,000	890
200	170	7,500	1,110
500	270	10,000	1,300
700	320	15,000	1,640
1,000	390	20,000	1,940
2,500	620	27,900	2,390

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 126,100 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 4,150. Consequently, the 68-percent confidence interval is from 121,950 to 130,250 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the inverval from 119,460 to 132,740 housing units with 90 percent confidence; and that the average estimate lies within the interval from 117,800 to 134,400 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 126,100 specified owner-occupied housing units with two bedrooms, 27,800 or 22.0 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 27,800 is approximately 2,130. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 22.0 percent is approximately 1.5 percentage

$$1.5 = (100)\left(\frac{27{,}800}{126{,}100}\right)\sqrt{\left(\frac{2{,}130}{27{,}800}\right)^2-\left(\frac{4{,}150}{126{,}100}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 20.5 to 23.5 percent; the 90-percent confidence interval is from 19.6 to 24.4 percent; and the 95-percent confidence interval is from 19.0 to 25.0 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 30,700 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999.

Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 2,900. The standard error of 27,800 is 2,130, as shown above. Table I also shows the standard error on an estimate of 30,700 to be approximately 2,220. Therefore, the standard error of the estimated difference of 2,900 is about:

$$3{,}080 = \sqrt{(2{,}130)^2 + (2{,}220)^2}$$

Consequently, the 68-percent confidence interval for the 2,900 difference is from −180 to 5,980 housing units. Therfore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from −2,030 to 7,830 housing units, and the 95-percent confidence interval is from −3,260 to 9,060. Thus, we cannot conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $10,000 and $14,999 is greater than the number of units valued between $15,000 and $19,999 since the 95-percent confidence interval includes zero or negative values. We also cannot conclude with either 90 percent or 68 percent confidence that the 30,700 is greater than the 27,800 since both the 90-percent and 68-percent confidence intervals includes zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $17,200 in 1974. The base of the distribution from which this median was determined is 126,100 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 126,100 is 1.8 percentage points:

$$1.8 = (100)\left(\frac{63{,}050}{126{,}100}\right)\sqrt{\left(\frac{3{,}100}{63{,}050}\right)^2 - \left(\frac{4{,}150}{126{,}100}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.4 to 53.6.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first three categories that 50,500 owner-occupied housing units with two bedrooms, or 40.4 percent, had a value less than $15,000; and an additional 27,800 owner-occupied housing units with two bedrooms, or 22.0 percent had a value between $15,000 and $19,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about:

$$\$15{,}000 + (\$5{,}000)\left(\frac{46.4 - 40.0}{22.0}\right) = \$16{,}500$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about:

$$\$15{,}000 + (\$5{,}000)\left(\frac{53.6 - 40.0}{22.0}\right) = \$18{,}100$$

Thus, the 95-percent confidence interval ranges from $16,500 to $18,100.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which

they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items.

The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 4,000 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that no new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Saginaw 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error table App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Saginaw, Mich., SMSA, 4,840 sample housing units were eligible for interview. Of this number, 110 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 270 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Saginaw, Mich., SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 110 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The quality or reliability of the independent estimate varied by SMSA, depending on

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957–1959 by the Census Bureau.

the completeness of the utility data or permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of total housing and an April 1970 independent estimate of total housing, generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period, 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the table below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory as well as estimates of characteristics of the 1970-1974 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in table I.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator

y = the denominator

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory and for Estimated Number of 1970-1974 Lost Units for the Saginaw, Mich., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	40	5,000	260
200	50	10,000	350
500	90	25,000	480
1,000	130	50,000	470
2,500	190		

Illustration of the use of the standard error table.—Table 2 in part C of this report shows that in this SMSA there were 10,500 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 350. Consequently, the 68-percent confidence interval is from 10,150 to 10,850 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 9,940 to 11,060 housing units with 90 percent confidence; and that the average estimate lies within the interval from 9,800 to 11,200 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 10,500 specified owner-occupied housing units with two bedrooms, 2,700, or 25.7 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 2,700 is approximately 200. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 25.7 percent is approximately 1.7 percentage points:

$$1.7 = (100)\left(\frac{2{,}700}{10{,}500}\right)\sqrt{\left(\frac{200}{2{,}700}\right)^2 - \left(\frac{350}{10{,}500}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 24.0 to 27.4 percent; the 90-percent confidence interval is from 23.0 to 28.4 percent; and the 95-percent confidence interval is from 22.3 to 29.1 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 1,400 specified owner-occupied units with two bedrooms valued between $20,000 and $24,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $20,000 and $24,999 and those valued between $15,000 and $19,999 is 1,300. The standard error of 2,700 is 200 as shown above. Table I also shows the standard error on an estimate of 1,400 to be approximately 150. Therefore, the standard error of the estimated difference of 1,300 is about

$$250 = \sqrt{(200)^2 + (150)^2}$$

Consequently, the 68-percent confidence interval for the 1,300 difference is from 1,050 to 1,550 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 900 to 1,700 housing units, and the 95-percent confidence interval is from 800 to 1,800. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $20,000 and $24,999 since the 95-percent con-

fidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $17,900 in 1974. The base of the distribution from which this median was determined is 10,500 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 10,500 is 1.8 percentage points:

$$1.8 = (100)\left(\frac{5,250}{10,500}\right)\sqrt{\left(\frac{260}{5,250}\right)^2 - \left(\frac{350}{10,500}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.4 to 53.6.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first three categories that 3,700 owner-occupied housing units with two bedrooms, or 35.2 percent, had a value less than $15,000; and an additional 2,700 owner-occupied housing units with two bedrooms, or 25.7 percent, had a value between $15,000 and $19,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{46.4 - 35.2}{25.7}\right) = \$17,200$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{53.6 - 35.2}{25.7}\right) = \$18,600$$

Thus, the 95-percent confidence interval ranges from $17,200 to $18,600.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on

the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census,** and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)

2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)

3. "In 1970, the definitional under-enumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)

4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)

5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:

a. heating fuel
b. renters paying extra for utilities and/or fuel
c. bathtub or shower
d. flush toilet
e. telephone availability
f. year structure built
g. value of home
h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These housholds were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories

had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 800 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 700 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Salt Lake City 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Salt Lake City, Utah, SMSA, 4,920 sample housing units were eligible for interview. Of this number, 120 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 280 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Salt Lake City, Utah, SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 120 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census, based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The quality or reliability of the independent estimate varied by SMSA, depending on the completeness of the

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

utility data or permit demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available, based on a comparison between the 1970 census estimate of total housing and an April 1970 independent estimate of total housing, generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibliity of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than 3 percent over the 10-year period, 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides

an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Salt Lake City, Utah SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	60	5,000	440
200	90	10,000	620
500	140	25,000	940
1,000	200	50,000	1,220
2,500	330	100,000	1,410

TABLE II. Standard Errors for Estimated Number of 1970-1974 Lost Units for the Salt Lake City, Utah SMSA

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	60	1,000	190
200	80	1,500	240
500	130	2,500	330
700	160		

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 26,800 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 960. Consequently, the 68-percent confidence interval is from 25,840 to 27,760 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 25,260 to 28,340 housing units with 90-percent confidence; and that the average estimate lies within the interval from 24,880 to 28,720 housing units with 95-percent confidence.

Table 2 in part C also shows that of the 26,800 specified owner-occupied housing units with two bedrooms, 5,700, or 21.3 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 5,700 is approximately 470. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 21.3 percent is approximately 1.6 percentage points:

$$1.6 = (100)\left(\frac{5{,}700}{26{,}800}\right)\sqrt{\left(\frac{470}{5{,}700}\right)^2 - \left(\frac{960}{26{,}800}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 19.7 to 22.9 percent; the 90-percent confidence interval is from 18.7 to 23.9 percent; and the 95-percent confidence interval is from 18.1 to 24.5 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 2,000 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 3,700. The standard error of 5,700 is 470 as shown above. Table I also shows the standard error on an estimate of 2,000 to be approximately 290. Therefore, the standard error of the estimated difference of 3,700 is about

$$550 = \sqrt{(470)^2 + (290)^2}$$

Consequently, the 68-percent confidence interval for the 3,700 difference is from 3,150 to 4,250 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68

percent of all possible samples. Similarly, the 90-percent confidence interval is from 2,820 to 4,580 housing units, and the 95-percent confidence interval is from 2,600 to 4,800. Thus, we can conclude with 95-percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $23,900 in 1974. The base of the distribution from which this median was determined is 26,800 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 26,800 is 1.8 percentage points:

$$1.8 = (100)\left(\frac{13,400}{26,800}\right)\sqrt{\left(\frac{690}{13,400}\right)^2 - \left(\frac{960}{26,800}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.4 to 53.6.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first four categories that 8,000 owner-occupied housing units with two bedrooms, or 29.9 percent, had a value less than $19,999 and an additional 6,900 owner-occupied housing units with two bedrooms, or 25.7 percent, had a value between $20,000 and $24,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$20,000 + (\$5,000)\left(\frac{46.4 - 29.9}{25.7}\right) = \$23,200$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$20,000 + (\$5,000)\left(\frac{53.6 - 29.9}{25.7}\right) = \$24,600$$

Thus, the 95-percent confidence interval ranges from $23,200 to $24,600.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional under-enumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 1,400 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 2,000 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Spokane 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Spokane, Wash., SMSA, 4,850 sample housing units were eligible for interview. Of this number, 120 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 390 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Spokane, Wash., SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000—$5,999 . . .		
$6,000—$9,999 . . .		
$10,000—$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 120 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The quality or reliability of the independent estimate varied by SMSA, depending on

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

the completeness of the utility data or permit demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of total housing and an April 1970 independent estimate of total housing, generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period, 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a re-

sult, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator

y = the denominator

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Spokane, Wash., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	50	5,000	370
200	70	10,000	450
500	110	25,000	660
1,000	150	50,000	790
2,500	240	100,000	480

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Spokane, Wash., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	40	1,500	180
200	60	2,500	240
500	100	3,500	300
700	120	4,000	320
1,000	150	4,700	350

Illustration of the use of the standard error tables.—Table 2 in part C of this report shows that in this SMSA there were 17,200 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 550. Consequently, the 68-percent confidence interval is from 16,650 to 17,750 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 16,320 to 18,080 housing units with 90 percent confidence; and that the average estimate lies within the interval from 16,100 to 18,300 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 17,200 specified owner-occupied housing units with two bedrooms, 5,500, or 32.0 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 5,500 is approximately 380. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 32.0 percent is approximately 1.6 percentage points:

$$2.0=(100)\left(\frac{5{,}500}{17{,}200}\right)\sqrt{\left(\frac{380}{5{,}500}\right)^2 - \left(\frac{550}{17{,}200}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 30.0 to 34.0 percent; the 90-percent confidence interval is from 28.8 to 35.2 percent; and the 95-percent confidence interval is from 28.0 to 36.0 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 4,400 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 1,100. The standard error of 5,500 is 380 as shown above. Table I also shows the standard error on an estimate of 4,400 to be approximately 340. Therefore, the standard error of the estimated difference of 1,100 is about

$$510=\sqrt{(380)^2 + (340)^2}$$

Consequently, the 68-percent confidence interval for the 1,100 difference is from 590 to 1,610 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 280 to 1,920 housing units, and the 95-percent confidence interval is from 80 to 2,120. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $16,900 in 1974. The base of the distribution from which this median was determined is 17,200 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 17,200 is 1.9 percentage points:

$$1.9=(100)\left(\frac{8,600}{17,200}\right)\sqrt{\left(\frac{430}{8,600}\right)^2-\left(\frac{550}{17,200}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.2 to 53.8.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first three categories that 6,500 owner-occupied housing units with two bedrooms, or 37.8 percent, had a value less than $15,000; and an additional 5,500, or 32.0 percent, had a value between $15,000 and $19,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{46.2-37.8}{32.0}\right) = \$16,300$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{53.8-37.8}{32.0}\right) = \$17,500$$

Thus, the 95-percent confidence interval ranges from $16,300 to $17,500.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one

living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)

2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)

3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)

4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)

5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:

a. heating fuel
b. renters paying extra for utilities and/or fuel
c. bathtub or shower
d. flush toilet
e. telephone availability
f. year structure built
g. value of home
h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items.

The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 900 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 1,400 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Tacoma 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error table App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Tacoma, Wash., SMSA, 4,880 sample housing units were eligible for interview. Of this number, 130 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 490 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Tacoma, Wash., SMSA was selected from three sample frames. Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000–$5,999 . . .		
$6,000–$9,999 . . .		
$10,000–$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties**, Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 130 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census, based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The quality or reliability of the independent estimate varied by SMSA, depending on the completeness of the utility data or

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

permit demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available, based on a comparison between the 1970 census estimate of total housing and an April 1970 independent estimate of total housing, generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period, 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude

of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory as well as estimates of characteristics of the 1970-1974 lost units. Linear interpolation should be used to determine the standard error for estimates not specifically shown in table I.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator

y = the denominator

σ_x = the standard error of the numerator

σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory and for Estimated Number of 1970-1974 Lost Units for the Tacoma, Wash., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	50	5,000	410
200	80	10,000	530
500	120	25,000	790
1,000	170	50,000	1,000
2,500	280	100,000	980

Illustration of the use of the standard error table.—Table 2 in part C of this report shows that in this SMSA there were 19,800 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 700. Consequently, the 68-percent confidence interval is from 19,100 to 20,500 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 18,680 to 20,920 housing units with 90 percent confidence; and that the average estimate lies within the interval from 18,400 to 21,200 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 19,800 specified owner-occupied housing units with two bedrooms, 6,600, or 33.3 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 6,600 is approximately 450. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 33.3 percent is approximately 1.9 percentage points:

$$1.9 = (100)\left(\frac{6{,}600}{19{,}800}\right)\sqrt{\left(\frac{450}{6{,}600}\right)^2 - \left(\frac{700}{19{,}800}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 31.4 to 35.2 percent; the 90-percent confidence interval is from 30.3 to 36.3 percent; and the 95-percent confidence interval is from 29.5 to 37.1 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 4,300 specified owner-occupied units with two bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 2,300. The standard error of 6,600 is 450 as shown above. Table I also shows the standard error on an estimate of 4,300 to be approximately 370. Therefore, the standard error of the estimated difference of 2,300 is about

$$580 = \sqrt{(450)^2 + (370)^2}$$

Consequently, the 68-percent confidence interval for the 2,300 difference is from 1,720 to 2,800 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 1,370 to 3,230 housing units, and the 95-percent confidence interval is from 1,140 to 3,460. Thus, we can conclude with 95-percent confidence that the number of 1974 owner-occupied housing

units with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $18,400 in 1974. The base of the distribution from which this median was determined is 19,800 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 19,800 is 2.0 percentage points:

$$2.0 = (100)\left(\frac{9,900}{19,800}\right)\sqrt{\left(\frac{530}{9,900}\right)^2 - \left(\frac{700}{19,800}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.0 to 54.0.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first three categories that 5,500 owner-occupied housing units with two bedrooms, or 27.8 percent, had a value less than $15,000; and an additional 6,600 owner-occupied housing units with two bedrooms, or 33.3 percent, had a value between $15,000 and $19,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{46.0 - 27.8}{33.3}\right) = \$17,700$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{54.0 - 27.8}{33.3}\right) = \$18,900$$

Thus, the 95-percent confidence interval ranges from $17,700 to $18,900.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.

2. Record check.—The comparison of census data with data obtained from an independent record source.

3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)

2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)

3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)

4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)

5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:

a. heating fuel
b. renters paying extra for utilities and/or fuel
c. bathtub or shower
d. flush toilet
e. telephone availability
f. year structure built
g. value of home
h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.

2. The correct number of housing units were interviewed at that address.

3. The correct information on "Year Built" was obtained.

4. The correct information on "Tenure" was obtained.

5. The correct information on "Household Composition" was obtained.

6. The correct information on "Type of Housing Unit" was obtained.

7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in

content reinterview."

2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 2,800 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 1,300 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimatior procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Washington 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-28
Ratio estimation procedure of the 1970 Census of Population and Housing App-28
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error tables App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-30
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Washington, D.C.-Md.-Va., SMSA, 14,300 sample housing units were eligible for interview. Of this number, 960 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 1,330 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Washington, D.C.-Md.-Va., SMSA was selected from two sample frames—units enumerated in the 1970 Census of Housing and Population (the 1970 census universe) and units constructed since the 1970 census (the new construction universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate used to select the AHS sample for the SMSA was chosen so that the desired designated sample size would result. The overall sampling rate for the SMSA did differ by central city and balance since the sample for this SMSA was split equally between the central city and the balance of the SMSA.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000–$5,999 . . .		
$6,000–$9,999 . . .		
$10,000–$14,999 .		
$15,000 and over . .		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for

either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the 1970 census universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 960 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

Within this SMSA, the factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the 1970 census universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), and one noninterview cell for new construction sample housing units.

The first-stage ratio estimation procedure was employed for all sample housing units from the 1970 census universe. This factor was computed separately for all sample housing units with each 1970 census universe noninterview cell mentioned above. The ratio estimation factor for each cell was equal to:

$$\frac{\text{1970 census count of housing units from 1970 census universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the 1970 census universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units al-

ready selected for other Census Bureau surveys were deleted from the 1970 census universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units	
New mobile homes	
"Other additions"	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" were applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with a more reliable estimate of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census, based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The reliability of this independent estimate varied by SMSA, depending on the completeness of the utility data or the permit-demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available based on a comparison between the 1970 census estimate of the April 1970 housing inventory and the April 1970 independent estimate of the housing inventory generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory for the Washington, D.C.-Md.-Va., SMSA, for the Central Cities, and for the Balance of the SMSA.**

(68 chances out of 100)

Size of estimate	Standard error		
	Total SMSA	In central cities	Not in central cities
500	210	140	240
1,000	300	200	330
2,500	480	320	530
5,000	680	450	750
10,000	960	630	1,050
25,000	1,500	960	1,650
50,000	2,090	1,290	2,290
100,000	2,890	1,610	3,120
250,000	4,180	920	4,350
500,000	4,890	-	4,420
750,000	4,390	-	1,400
1,000,000	1,870	-	-

TABLE II. **Standard Errors for Estimated Number of 1970-1974 Lost Units for the Washington, D.C.-Md.-Va., SMSA, for the Central Cities, and for the Balance of the SMSA**

(68 chances out of 100)

Size estimate	Standard error		
	Total SMSA	In central cities	Not in central cities
100	90	70	100
200	130	100	150
500	210	150	230
700	250	180	280
1,000	300	220	330
1,500	370	270	410
2,500	470	350	540
3,500	560	420	640
5,000	660	510	790
7,500	800	640	1,000
10,000	920	760	1,190
15,000	1,110	980	-
20,000	1,270	-	-
23,300	1,360	-	-

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory. Table II presents the standard errors for estimates of characteristics of the 1970 units lost between 1970-1974. Linear interpolation should be used to determine the standard error for estimates not specifically shown in tables I and II.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (–) to plus (+).

Illustration of the use of the standard error tables.—Table B-2 in part C of this report shows that in the central city of this SMSA there were 56,400 specified owner-occupied housing units with 3 or more bedrooms in 1974. Interpolation in column 2 of table I of this appendix shows that the standard error of an estimate of this size is approximately 1,330. Consequently, the 68-percent confidence interval is from 55,070 to 57,730 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with 3 or more bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 54,270 to 58,530 housing units with 90 percent confidence; and that the average estimate lies within the interval from 53,740 to 59,060 housing units with 95 percent confidence.

Table 3-2 in part C also shows that of the 56,400 specified owner-occupied housing units with 3 or more bedrooms, 4,600, or 8.2 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 4,600 is approximately 430. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding this illustration), the standard error of 8.2 percent is approximately 0.7 percentage points:

$$0.7=(100)\left(\frac{4{,}600}{56{,}400}\right)\sqrt{\left(\frac{430}{4{,}600}\right)^2-\left(\frac{1{,}330}{56{,}400}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 7.5 to 8.9 percent; the 90-percent confidence interval is from 7.1 to 9.3 percent; and the 95-percent confidence interval is from 6.8 to 9.6 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table B-2 in part C of this SMSA report shows that in 1974 there were 900 specified owner-occupied units with 3 or more bedrooms valued between $10,000 and $14,999. Thus the apparent difference between the number of 1974 owner-occupied units with 3 or more bedrooms valued between $10,000 and $14,999 and those valued between $15,000 and $19,999 is 3,700. The standard error of 4,600 is 430, as shown above. Table I also shows the standard error on an estimate of 900 to be approximately 190. Therefore, the standard error of the estimated difference of 3,700 is about

$$470=\sqrt{(430)^2+(190)^2}$$

Consequently, the 68-percent confidence interval for the 3,700 difference is from 3,230 to 4,170 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 2,950 to 4,450 housing units, and the 95-percent confidence interval is from 2,760 to 4,640. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied housing units with 3 or more bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $10,000 and $14,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I or II in conjunction with the formula for the standard

error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table B-2 in part C of this report shows the median value of owner-occupied housing units with 3 or more bedrooms in the central city of this SMSA was $33,700 in 1974. The base of the distribution from which this median was determined is 56,400 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 56,400 is 1.3 percentage points:

$$1.3=(100)\left(\frac{28,200}{56,400}\right)\sqrt{\left(\frac{1,000}{28,200}\right)^2-\left(\frac{1,330}{56,400}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 47.4 to 52.6.

3. From table B-2 in part C of this report it can be seen by cumulating the frequencies for the first five categories that 13,100 owner-occupied housing units with 3 or more bedrooms, or 23.2 percent, had a value less than $25,000; and an additional 17,300, or 30.7 percent, had a value between $25,000 and $34,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$25,000+(\$35,000-\$25,000)\left(\frac{47.4-23.2}{30.7}\right)=\$32,900.$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$25,000+(\$35,000-\$25,000)\left(\frac{52.6-23.2}{30.7}\right)=\$34,600.$$

Thus, the 95-percent confidence interval ranges from $32,900 to $34,600.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.

2. Record check.—The comparison of census data with data obtained from an independent record source.

3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing**

Characteristics as Measured by Reinterviews. Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional under-enumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."

2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which were expected to have relatively high inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time

of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 11,300 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that approximately 700 new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.

Wichita 1974

APPENDIX B—Source and Reliability of the Estimates

SAMPLE DESIGN App-26
Selection of the sample App-26
Building loss sample selection App-27
1970 Census of Population and Housing App-27
ESTIMATION App-27
1974 housing inventory App-27
1970-1974 lost units App-29
Ratio estimation procedure of the 1970 Census of Population and Housing App-29
RELIABILITY OF THE ESTIMATES App-29
Sampling errors for the AHS-SMSA sample App-29
Illustration of the use of the standard error table App-30
Differences App-30
Illustration of the computation of the standard error of a difference App-30
Medians App-31
Illustration of the computation of the standard error of a median App-31
Nonsampling errors App-31
1970 census App-31
AHS-SMSA App-32

SAMPLE DESIGN

The estimates for each of the 19 SMSA's are based on data collected from the 1974 Annual Housing Survey (AHS), which was conducted by the Bureau of the Census, acting as collection agent for the Department of Housing and Urban Development. In each of the 19 SMSA's, the data were collected during a 12-month period from April 1974 through March 1975 with one-twelfth of the sample units being visited each month.

Data for a second group of 21 SMSA's were collected for the AHS from April 1975 through March 1976, and for a third group of 20 SMSA's, from April 1976 through March 1977. The sample housing units for each group of SMSA's will be enumerated every three years on a rotating basis.

The four largest SMSA's in each group are represented separately by a sample of 15,000 designated housing units evenly divided between the central city and the balance of the respective SMSA. All remaining SMSA's are each represented by a sample of 5,000 designated housing units distributed proportionately between the central city and the balance of the respective SMSA based on the distribution of total housing units in each sector.

The four largest SMSA's in the first group (Year I SMSA's) are: Boston, Mass., Detroit, Mich., Los Angeles-Long Beach, Calif., and Washington, D.C.-Md.-Va.

The remaining SMSA's in this first group are: Albany-Schenectady-Troy, N.Y., Anaheim-Santa Ana-Garden Grove, Calif., Dallas, Tex., Fort Worth, Tex., Memphis, Tenn.-Ark., Minneapolis-St. Paul, Minn., Newark, N.J., Orlando, Fla., Phoenix, Ariz., Pittsburgh, Pa., Saginaw, Mich., Salt Lake City, Utah, Spokane, Wash., Tacoma, Wash., and Wichita, Kans.

In the Wichita, Kans., SMSA, 4,930 sample housing units were eligible for interview. Of this number, 140 interviews were not obtained because, for occupied housing units, the occupants were not at home after repeated visits or were unavailable for some other reason; or, for vacant units, no informed respondent could be found after repeated visits. In addition to units eligible for interview, 390 units were visited and found to be condemned, unfit, demolished, converted to group quarters use, etc., and, therefore, were not eligible for interview.

Selection of the sample.—The AHS sample for the Wichita, Kans., SMSA was selected from three sample frames: Housing units enumerated in the 1970 Census of Housing and Population in areas under the jurisdiction of permit-issuing offices for this SMSA (the permit-issuing universe), units constructed since the 1970 census in permit-issuing areas (the new construction universe), and units located in areas not under the jurisdiction of permit-issuing offices (the nonpermit universe). Sampling operations, described in the following paragraphs, were performed separately within the central city and the balance of this SMSA for each of these sample frames. The overall sampling rate was about the same for the sample selected from both the central city and the balance of this SMSA since the sample was distributed proportionately between the central city and the balance of this SMSA according to the distribution of total units in each sector.

The major portion of the sample was selected from a file which represented the 20-percent sample of units enumerated in the 1970 Census of Housing and Population. This file contained records for occupied housing units, vacant housing units, and units in certain special places or group quarters. Sampling operations were done separately for the special place and group quarters records and for the occupied and vacant housing unit records. Before the sample was selected from records of the occupied and vacant units, the occupied housing unit records were stratified by race of head (non-Negro, Negro), and the vacant housing unit records were stratified into four categories pertaining to the value or rent associated with the vacant housing units. The occupied housing unit records were further stratified so that each unit was assigned to one of 50 strata according to its tenure, family size, and household income categories as illustrated by the following table.

Household income	TENURE	
	Owner Family size	Renter Family size
	1 2 3 4 5+	1 2 3 4 5+
Under $3,000		
$3,000–$5,999 ...		
$6,000–$9,999 ...		
$10,000–$14,999 .		
$15,000 and over ..		

Thus, for this SMSA, the occupied housing unit records from this universe were assigned to one of 100 strata for either the central city or for the balance, and the vacant housing unit records were assigned to one of the four vacant strata for either the central city, or for the balance of the SMSA. A sample of housing unit records was then selected at a sampling rate twice that which had been determined necessary to produce the correct sample size. The housing unit record adjacent to each sample housing unit record was also selected to be in sample thereby insuring the necessary designated sample size.

Before the sample was selected from the group quarters and special place records, the records were stratified by census tract and census enumeration district (ED) within the central city and within the balance of the SMSA. A sample of special place records was then selected at a sampling rate four times that which had been determined necessary to produce the correct sample size. However, at the time of enumeration the units at each of the sample special places were listed and subsampled at a rate that would produce an expected four sample units, thereby insuring the necessary designated sample size.

The second frame from which this SMSA sample was selected was a list of new construction building permits issued since 1970 (i.e., the new construction universe). The sample selection was an independent operation within this SMSA. Prior to sample selection, the list of permits was chronologically stratified by the date the permits were issued and clusters of an expected four (usually adjacent) housing units were formed. These clusters were then sampled for inclusion at the overall sampling rate.

The remainder of the AHS sample was selected from a frame consisting of a list of areas not under the jurisdiction of permit-issuing offices (i.e., the nonpermit universe). The first step in the sampling operation for this universe was the selection, using the overall sampling rate, of a sample of census enumeration districts within these areas. Prior to this sample selection, the ED's were stratified by census tract within the central city and within the balance of the SMSA. The probability of selection for an ED was proportionate to the following measure of size:

$$\frac{\text{Number of housing units in 1970 census ED} + \dfrac{\text{Group quarters population in 1970 census ED}}{3}}{4}$$

The sample ED's were then divided into segments; i.e., small land areas with well-defined boundaries having an expected size of four (usually adjacent) housing units. Those segments, with an expected size which was a multiple of four, were further subdivided at the time of enumeration to produce an expected four sample housing units.

The next step was the selection of one of these segments within each sample ED. In the sample segments, all units in existence at the time of interview are in sample. Thus, units enumerated in the 1970 census as well as units built since the 1970 census are included.

Building loss sample selection.—Statistics in this report provide estimates of the 1970 characteristics of housing units removed from the inventory since 1970 (i.e., units enumerated in the 1970 census that no longer existed as part of the housing inventory at the time of the AHS enumeration). For the permit-issuing universe, the sample upon which these estimates were based was the regular AHS sample described above. Since the addresses for these sample units were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, estimates were obtained for both whole structure losses (i.e., lost units in structures in which all housing units were removed from the inventory) and part structure losses (i.e., lost units in structures in which some, but not all, of the units were removed from the inventory).

For the nonpermit universe, it was necessary to select an independent sample with which to measure lost units because of the area sample technique employed for the regular AHS sample. For this independent sample, a cluster of four (usually adjacent) housing unit addresses was selected from the 1970 census listing of addresses for each sample ED. Since these addresses were known to exist in 1970, the AHS interviewer determined those sample units that were no longer considered part of the housing inventory. From these units, only estimates of whole structure losses were obtained (i.e., lost units in structures in which all housing units were removed from the inventory).

1970 Census of Population and Housing.—The estimates pertaining to the 1970 housing inventory (i.e., the housing inventory that existed at the time of the 1970 census) are based on either 20-percent, 15-percent, or 5-percent sample data collected in April 1970 for the Decennial Census of Population and Housing. A detailed description of the sample design employed for the 1970 census can be obtained in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties, Part 1.**

ESTIMATION

The AHS-SMSA sample produced estimates of two types: Estimates pertaining to housing inventory characteristics at enumeration time, and estimates pertaining to characteristics of units removed from the housing inventory since 1970. Each type of estimate employed a separate, though similar, estimation procedure as described below.

1974 housing inventory.—The AHS estimations of characteristics of the 1974 housing inventory employed a two-stage ratio estimation procedure. However, prior to implementation of this procedure, the basic weight (i.e., the inverse of the probability of selection) for each interviewed sample housing unit was adjusted to account for the 140 noninterviews previously mentioned. The noninterview adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of interviewed housing units + Weighted count of noninterviewed housing units}}{\text{Weighted count of interviewed housing units}}$$

The factor was computed separately for sample housing units within the central city and within the balance of the SMSA. Within each sector a noninterview factor was computed separately for 54 noninterview cells for sample housing units from the permit-issuing universe (where the cells consisted of one or more of the different strata used in the stratification of the universe as previously illustrated), one noninterview cell for new construction sample housing units, and one noninterview cell for sample units from the nonpermit universe. Sample housing units from the nonpermit universe, identified as being built April 1, 1970, or later, were considered as new construction units for the purpose of the noninterview adjustment.

The first-stage ratio estimation procedure was employed for all sample housing units from the permit-issuing universe. This factor was computed separately for all sample housing units within each permit-issuing universe noninterview cell mentioned above. The ratio estimation factor for each cell was as follows:

$$\frac{\text{1970 census count of housing units from permit-issuing universe in a cell}}{\text{AHS sample estimate of 1970 housing units from the cell}}$$

The numerators of the ratios were obtained from the 1970 Census of Population and Housing 20-percent file of units enumerated in areas under the jurisdiction of permit-issuing offices. The denominators of the ratios were obtained from weighted estimates of all AHS sample units within the corresponding ratio estimation categories using the existing weight (i.e., the basic weight times the noninterview factor). The computed first-stage ratio estimation factor was then applied to the existing weight for each sample unit within the corresponding first-stage ratio estimation category.

The effect of this ratio estimation procedure was to reduce somewhat the variation in sample size for strata used in the sample selection for the permit-issuing universe. Ordinarily, this would have been controlled by sampling within the strata during the sample selection process. However, prior to the AHS sample selection within each SMSA, units already selected for other Census Bureau surveys were deleted from the permit-issuing universe. Thus, some variation in sample size was introduced during the AHS sample selection process.

The second-stage ratio estimation procedure was employed for all sample units that were conventional new construction units (i.e., conventional units built after April 1970), new mobile homes (i.e., mobile homes placed after April 1970), or "other addition" units (i.e., units added by conversion of 1970 units or from other sources). This procedure was designed to adjust the AHS sample estimates of such units to independently derived current estimates available for these types of units. This adjustment was necessary to correct for known deficiencies in the AHS sample with regard to representation of these units (see the section on nonsampling error).

The second-stage ratio estimation factors were computed separately for each of the cells in the table below using the following formula:

$$\frac{\text{October 1974 independent estimate of housing units in that cell for the SMSA}}{\text{AHS sample estimate of housing units in that cell for the SMSA}}$$

Second-Stage Ratio Estimation Cells

Conventional new construction units.	
New mobile homes	
"Other additions".	

The numerators of the ratios were derived by applying the following factors: (1) For the "conventional new construction units" cell, a national trend for missed conventional new construction was applied to the 1968 and 1969 building permits issued in this SMSA; (2) for the "new mobile homes" cell, a 1970 census relationship between total new construction and new mobile homes that existed in this SMSA for the 1965-1970 period was applied; and (3) for the "other additions" cell, rates from SCARF[1] for "other additions" was applied.

The denominators of the ratios were obtained from the weighted estimates for the AHS sample units within each cell, using the existing weight after the first-stage ratio estimation procedure. The computed second-stage ratio estimation factor was then applied to the existing weight for each sample unit in the corresponding second-stage ratio estimation category.

The effect of the second-stage ratio estimation procedure, as well as the overall estimation procedure, was to reduce the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from that for the SMSA as a whole, it can be expected that the sample estimates will be improved when the sample housing population is brought into agreement with known estimates of the SMSA housing population.

In some of the Year I SMSA's, a third-stage ratio estimation procedure was also employed. This procedure involved the ratio estimation of the AHS weighted sample estimate of the October 1974 housing inventory to an independent estimate of the SMSA's October 1974 housing inventory. This estimate was derived by using the 1970 census estimate of the April 1970 housing inventory in conjunction with an estimate of change in the housing inventory since the census, based on either administrative records from utility companies (where available) or estimates of new construction permits and post-census demolition data. The quality or reliability of the independent estimate varied by SMSA, depending on

[1] SCARF denotes the Survey of Components of Change and Residential Finance, a survey conducted in 1957-1959 by the Census Bureau.

the completeness of the utility data or permit demolition data. However, a measure of the reliability (i.e., the relative bias) of these independent estimates was available, based on a comparison between the 1970 census estimate of total housing and an April 1970 independent estimate of total housing, generated from the procedure used to produce the October 1974 independent estimates. As a consequence, the independent estimates were only employed in those SMSA's where the estimated relative bias was low enough (i.e., 3 percent or less over a 10-year period) to preclude the possibility of doing more harm than good to the survey results.

For this SMSA, this independent estimate was not employed because of the lack of reliability of the estimate (i.e., the bias was more than three percent over the 10-year period, 1960-1970).

1970-1974 lost units.—The estimate of characteristics of the 1970-1974 lost units employed a one-stage ratio estimation procedure similar to the first-stage ratio estimation procedure mentioned above. The 1970-1974 lost units file was matched to the 1970 census file to obtain the 1970 characteristics of the lost units. As a result, some lost units did not match. Thus, following the implementation of the ratio estimation procedure, the weight for each matched lost unit was adjusted to account for the nonmatched lost units. The nonmatch adjustment factor was equal to the following ratio:

$$\frac{\text{Weighted count of matched lost units} + \text{Weighted count of nonmatched lost units}}{\text{Weighted count of matched lost units}}$$

Within each sector of the SMSA, this factor was computed separately for whole-structure losses and part-structure losses. The effect of this estimation procedure was to reduce the sampling error below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection.

Ratio estimation procedure of the 1970 Census of Population and Housing.—This report presents data on the housing characteristics of the 1970 housing inventory from the 1970 Census of Population and Housing. The statistics based on 1970 census sample data employed a ratio estimation procedure which was applied separately for each of the three census samples. A detailed description of the ratio estimation procedure employed for the 1970 census can be found in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

RELIABILITY OF THE ESTIMATES

There are two types of possible errors associated with estimates based on data from sample surveys—sampling and nonsampling errors. The "accuracy" of a survey estimate is determined by the joint effects of the sampling and nonsampling errors. The following is a description of the sampling and nonsampling errors associated with the AHS-SMSA sample and of the nonsampling errors associated with the 1970 census estimates. A description of the sampling errors associated with the sample estimates from the 1970 census is in the 1970 Census of Housing report, Volume I, **Housing Characteristics for States, Cities, and Counties,** Part 1.

Sampling errors for the AHS-SMSA sample.—The particular sample used for this survey is one of a large number of possible samples of the same size that could have been selected using the same sample design. Even if the same schedules, instructions, and enumerators were used, estimates from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is defined as the sampling error. The standard error of a survey estimate attempts to measure this variation among the estimates from the possible samples and thus is a measure of the precision with which an estimate from a sample approximates the average result of all possible samples.

As calculated for this report, the standard error also partially measures the variation in the estimates due to response and enumerator errors (nonsampling errors), but it does not measure, as such, any systematic biases in the data. Therefore, the accuracy of the estimates depends on both the sampling and nonsampling errors, measured by the standard error, biases, and some additional nonsampling errors not measured by the standard error.

The sample estimate and its estimated standard error enable one to construct interval estimates in which the interval includes the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these surveyed under essentially the same general conditions, and an estimate and its estimated standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average result of all possible samples may or may not be contained in any particular computed interval. For a particular sample, however, one can say with specified confidence that the average result of all possible samples is included in the constructed interval.

The figures presented in the tables below are approximations to the standard errors of various estimates shown in this SMSA report. In order to derive standard errors that would be applicable to a wide

variety of items and also could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than precise standard errors for any specific item.

Table I presents the standard errors applicable to estimates of characteristics of the 1974 housing inventory as well as estimates of characteristics of the 1970-1974 lost units. Linear interpolation should be used to determine the standard errors for estimates not specifically shown in table I.

The reliability of an estimated percentage depends upon the size of the percentage and the size of the total upon which the percentage is based. An approximation to the standard error of a percentage may be obtained by using the following formula:

Let x = the numerator
y = the denominator
σ_x = the standard error of the numerator
σ_y = the standard error of the denominator

The standard error of the percentage (i.e., (100) (x/y)) is approximately equal to

$$(100)\,(x/y)\sqrt{\left(\frac{\sigma_x}{x}\right)^2 - \left(\frac{\sigma_y}{y}\right)^2}$$

The standard errors of x and y should be obtained from the appropriate standard error tables. For ratios, where x is not a subclass of y, the above formula underestimates the standard error of the ratio when there is little or no correlation between x and y. For this type of ratio a better approximation of the standard error may be obtained by changing the sign in the formula from minus (—) to plus (+).

TABLE I. **Standard Errors for Estimated Number of Housing Units in the 1974 Housing Inventory and for Estimated Number of 1970-1974 Lost Units for the Wichita, Kans., SMSA**

(68 chances out of 100)

Size of estimate	Standard error	Size of estimate	Standard error
100	50	5,000	380
200	70	10,000	510
500	120	25,000	760
1,000	170	50,000	950
2,500	260	100,000	890

Illustration of the use of the standard error table.—Table 2 in part C of this report shows that in this SMSA there were 26,700 specified owner-occupied housing units with two bedrooms in 1974. Interpolation in table I of this appendix shows that the standard error of an estimate of this size is approximately 770. Consequently, the 68-percent confidence interval is from 25,930 to 27,470 housing units. Therefore, a conclusion that the average estimate of 1974 specified owner-occupied housing units with two bedrooms, derived from all possible samples, lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, we could conclude that the average estimate derived from all possible samples lies within the interval from 25,470 to 27,930 housing units with 90 percent confidence; and that the average estimate lies within the interval from 25,160 to 28,240 housing units with 95 percent confidence.

Table 2 in part C also shows that of the 26,700 specified owner-occupied housing units with two bedrooms, 7,800, or 29.2 percent, were valued between $15,000 and $19,999. Table I of this appendix shows that the standard error for 7,800 is approximately 450. Therefore, using the formula for the standard error of a percentage (as shown in the paragraph preceding table I of this appendix), the standard error of 29.2 percent is approximately 1.5 percentage points:

$$1.5 = (100)\left(\frac{7{,}800}{26{,}700}\right)\sqrt{\left(\frac{450}{7{,}800}\right)^2 - \left(\frac{700}{26{,}700}\right)^2}$$

Consequently, the 68-percent confidence interval, as shown by these data, is from 27.7 to 30.7 percent; the 90-percent confidence interval is from 26.8 to 31.6 percent; and the 95-percent confidence interval is from 26.2 to 32.2 percent.

Differences.—The standard errors shown are not directly applicable to differences between two estimates. The standard error of a difference between estimates is approximately equal to the square root of the sum of the squares of the standard errors of each estimate considered separately. This formula is quite accurate for the difference between estimates of the same characteristic in two different SMSA's or the difference between separate and uncorrelated characteristics in the same SMSA. If, however, there is a high positive correlation between the two characteristics, the formula will overestimate the true error.

Illustration of the computation of the standard error of a difference.—Table 2 in part C of this SMSA report shows that in 1974 there were 3,100 specified owner-occupied units with two bedrooms valued between $20,000 and $24,999. Thus the apparent difference between the number of 1974 owner-occupied units with two bedrooms valued between $20,000 and $24,999 and those valued between $15,000 and $19,999 is 4,700. The standard error of 7,800 is 450 as shown above. Table I also shows the standard error on an estimate of 3,100 to be approximately 290. Therefore, the standard error of the estimated difference of 4,700 is about

$$540 = \sqrt{(450)^2 + (290)^2}$$

Consequently, the 68-percent confidence interval for the 4,700 difference is from 4,160 to 5,240 housing units. Therefore, a conclusion that the average estimate, derived from all possible samples, of this difference lies within a range computed in this way would be correct for roughly 68 percent of all possible samples. Similarly, the 90-percent confidence interval is from 3,840 to 5,560 housing units, and the 95 percent confidence interval is from 3,620 to 5,780. Thus, we can conclude with 95 percent confidence that the number of 1974 owner-occupied

housing unis with two bedrooms, valued between $15,000 and $19,999 is greater than the number of units valued between $20,000 and $24,999 since the 95-percent confidence interval does not include zero or negative values.

Medians.—For the medians presented in certain tables, the sampling error depends on the size of the base and on the distribution upon which the median is based. An approximate method for measuring the reliability of the estimated median is to determine an interval about the estimated median such that there is a stated degree of confidence that the average median from all possible samples lies within the interval. The following procedure may be used to estimate confidence limits of a median based on sample data:

(1) From table I in conjunction with the formula for the standard error of a percentage, determine the standard error of a 50-percent characteristic on the base of the median;

(2) add to and subtract from 50 percent, the standard error determined in step 1; and

(3) using the distribution of the characteristic, read off the confidence interval corresponding to the two points established in step 2.

For about 68 out of 100 possible samples, the average median from all possible samples would lie between these two values.

A two-standard-error confidence interval may be determined by finding the values corresponding to 50 percent plus and minus twice the standard error determined in step 1. For about 95 out of 100 possible samples, the average median from all possible samples would lie between these two values.

Illustration of the computation of the standard error of a median.—Table 2 in part C of this report shows the median value of owner-occupied housing units with two bedrooms was $16,100 in 1974. The base of the distribution from which this median was determined is 26,700 housing units.

1. Table I, in conjunction with the formula for the standard error of a percentage, shows that the standard error of 50 percent on a base of 26,700 is 1.6 percentage points:

$$1.6 = (100)\left(\frac{13,350}{26,700}\right)\sqrt{\left(\frac{570}{13,350}\right)^2 - \left(\frac{770}{26,700}\right)^2}$$

2. To obtain a two-standard-error confidence interval on the estimated median, add to and subtract from 50 percent twice the standard error determined in step 1. This yields percentage limits of 46.8 to 53.2.

3. From table 2 in part C of this report it can be seen by cumulating the frequencies for the first three categories that 11,500 owner-occupied housing units with two bedrooms, or 43.1 percent, had a value less than $15,000; and an additional 7,800 owner-occupied housing units with two bedrooms, or 29.2 percent, had a value between $15,000 and $19,999.

By linear interpolation, the lower limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{46.8 - 43.1}{29.2}\right) = \$15,600$$

Similarly, the upper limit of the 95-percent confidence interval is found to be about

$$\$15,000 + (\$5,000)\left(\frac{53.2 - 43.1}{29.2}\right) = \$16,700$$

Thus, the 95-percent confidence interval ranges from $15,600 to $16,700.

Nonsampling errors.—In general, nonsampling errors can be attributed to many sources: Inability to obtain information about all cases, definitional difficulties, differences in the interpretation of questions, inability or unwillingness to provide correct information on the part of respondents, mistakes in recording or coding the data, and other errors of collection, response, processing coverage, and estimation for missing data. As can be seen from the above list, nonsampling errors are not unique to sample surveys since they can, and do, occur in complete censuses as well.

Obtaining a measurement of the total nonsampling error associated with the estimates from a survey is very difficult, considering the number of possible sources of error. However, an attempt was made to measure some of the nonsampling errors associated with the estimates for both the 1970 Census of Population and Housing and the 1974 AHS-SMSA sample.

1970 census.—Several studies were conducted to measure two types of general errors—"coverage" errors and "content" errors—associated with 1970 census estimates.

The "coverage" errors determined how completely housing units were counted in the census and included the following:

1. Space errors.—Errors in which both the living quarters and its occupants were missed in the census or in which they were counted more than once (overenumerated). Space errors usually are the largest component of housing coverage error.

2. Definitional errors.—This type of error is best described by an example. Consider an address that appears in the census listings as a single-family home and consequently receives only one census questionnaire. The home is owned by a person who has converted part of the house into a separate apartment for use by another family. Since only one questionnaire was received by the owner, he might list the other family as members of his household. In this case, only one living quarters would have been counted where two existed.

3. Occupancy errors.—Errors of incorrect occupancy classification for enumerated units; i.e., vacant units that are improperly enumerated as occupied and vice versa.

The "content" error measured the accuracy of the data collected for enumerated housing units. Studies associated with the measurement of the content error measured the extent of errors arising from the erroneous or unreliable reporting of housing characteristics on the census questionnaire. In these studies, content errors were measured by the following methods:

1. Reinterview.—Households originally enumerated in the census were revisited and a second observation was obtained. These two readings were assumed to be independent.
2. Record check.—The comparison of census data with data obtained from an independent record source.
3. Comparison of census data with that obtained from other sample surveys.

The detailed results of these studies, as well as the methodology employed, can be obtained in the 1970 Census of Population and Housing, Evaluation and Research Program Reports, Series PHC(E)-5, **The Coverage of Housing in the 1970 Census**, and Series PHC(E)-10, **Accuracy of Data for Selected Housing Characteristics as Measured by Reinterviews.** Some of the results are presented for the country as a whole below:

1. "The total missed rate for housing units in 1970 is estimated to be 2.5 percent"; i.e., for each 100 units that were finally enumerated in the census, an estimated 2.5 were missed. (PHC(E)-5, p.3)
2. "The occupied space missed rate for the total United States in 1970 is estimated at 1.7 percent." About one-fourth of the errors occurred within structures included in the census and about three-fourths were due to missed structures. (PHC(E)-5, p.4)
3. "In 1970, the definitional underenumeration rate was 0.3 of 1 percent and the overenumeration rate was 0.04 of 1 percent for a total error of 0.34 of 1 percent." (PHC(E)-5, p.11)
4. "Most of the vacant units that were enumerated as occupied were procedure errors, whereas most occupied units misclassified as vacant were caused by enumerator error." (PHC(E)-5, p. 16) Approximately 16.5 percent of all units initially enumerated as vacant should have been enumerated as occupied, and about 0.3 of 1 percent of all units initially enumerated as occupied should have been enumerated as vacant. (PHC(E)-5, p. 15)
5. "Generally, owners reported housing data more consistently than renters; responses for occupied units were more consistent than those for vacant units; and respondents in single-unit structures reported more consistently than those in multi-unit structures." (PHC(E)-10, p. 6) The above statement applied to the following items:
   a. heating fuel
   b. renters paying extra for utilities and/or fuel
   c. bathtub or shower
   d. flush toilet
   e. telephone availability
   f. year structure built
   g. value of home
   h. seasonal vacancy status

"Heating fuel, year structure built, and value of home are more consistently reported for units with a non-Negro head, while bathtub or shower and flush toilet are more consistently reported for units with a Negro head." (PHC(E)-10, p. 8)

The results of these studies were based on sample data so there is sampling error associated with these estimates of nonsampling error. The possibility of such errors should be taken into account when considering the results of these studies. A detailed description of the sample design and estimation procedure for each study is given in the publications mentioned above.

AHS-SMSA.—For the 1974 AHS-SMSA sample, a study was conducted to obtain a measurement of some of the components of the nonsampling error associated with the AHS estimates. A reinterview program was conducted for a sample of the AHS households. These households were revisited and answers to some of the questions on the AHS questionnaire were obtained again. The original interview and the reinterview were assumed to be two independent readings and, thus, were the basis for the measurement of the accuracy of the AHS data collected from enumerated households.

As part of the reinterview, a check was made at each of these households to determine if the following was done during the original interview:

1. The correct unit was visited.
2. The correct number of housing units were interviewed at that address.
3. The correct information on "Year Built" was obtained.
4. The correct information on "Tenure" was obtained.
5. The correct information on "Household Composition" was obtained.
6. The correct information on "Type of Housing Unit" was obtained.
7. The correct information on "Occupancy Status" was obtained.

This check was made for interviewer evaluation and control. That is, tolerance limits were derived to determine which interviewers passed or failed this reinterview with regard to the above items. The results of this study are presented in the following census memorandum, "Reinterview Results for Annual Housing Survey-SMSA Sample; 1974."

Some of the results of this study are presented below. (Please note that these results are based on the reinterviews across all Year I SMSA's and not specifically for this SMSA.)

1. "The results indicate that the interviewers are doing a good job. From a total of 230 interviewers checked, 9 failed in coverage reinterview, 2 in household composition, and 6 in content reinterview."
2. A moderate level of inconsistency in responses existed between the original interview and the reinterview for most of the items selected for the reinterview. The items with the higher levels of inconsistency tended to be the attitude and opinion items which

were expected to have higher inconsistency levels.

3. "Our bias indicator, the net difference rate, revealed 7 categories out of 78 (78 from a total of 87 categories had enough data to compute reliable measures of response error) were significantly different from zero."

The results of this study were based on sample data so there is sampling error associated with these estimates of nonsampling error. Therefore, the possibility of such errors should be taken into account when considering the results.

With respect to errors of coverage and estimation for missing data, it is believed that the AHS new construction sample had deficiencies with regard to the representation of both conventional new construction and new mobile homes in permit-issuing areas. During the sampling of building permits, only those issued January 1, 1970, or later were eligible to be sampled to represent conventional new construction in permit-issuing areas for this SMSA. It had been assumed that units with permits issued prior to 1970 would have been completed by the time of the 1970 census (i.e., April 1970) and, therefore, would have been represented in the sample selected from 1970 census units. However, it has been estimated that, nationally, there were about 600,000 conventional new construction units built after April 1970 whose permits were issued prior to January 1970 (most of which were issued in 1968 and 1969). Although it is not known exactly, an estimated 1,000 conventional new construction units in this SMSA had permits issued prior to January 1970 and, therefore, were missed by the 1974 AHS-SMSA survey. This estimate was obtained by applying the national relationship, between the number of units authorized by 1968 and 1969 permits and the number of those units built after April 1970, to the number of units authorized by the 1968 and 1969 permits for this SMSA.

In addition, unlike the procedure for conventional new construction, there is no sampling procedure specifically for new mobile homes (and trailers) in permit-issuing areas. However, new mobile homes in permit-issuing areas do come into sample if the mobile homes are located in mobile home parks, identified as such in the 1970 census. Nonetheless, new mobile homes in permit-issuing areas that are located in mobile home parks, not in existence at the time of the 1970 census or not identified as such in the 1970 census, have no chance of coming into the AHS sample. Although it is not known exactly, it is estimated that no new mobile homes in permit-issuing areas were missed by the 1974 AHS-SMSA survey in this SMSA. This estimate was obtained by applying the relationship for this SMSA, between total units built between 1965 and 1970, and mobile homes built between 1965 and 1970, to the AHS estimate of total units built since April 1970. The second-stage ratio estimation procedure was employed to reduce the effect of both these deficiencies, although some bias in the AHS sample still exists.

With respect to errors associated with processing, the rounding of estimates introduces another source of error in the data, the severity of which depends on the statistic being measured. The effect of rounding is significant relative to the sampling error only for small percentages, median number of persons, and median number of rooms when these figures are derived from relatively large bases. This means that confidence intervals formed from the standard errors given may be distorted, and this should be taken into account when considering the results of the survey.